

# Averaged Iterative Water-Filling Algorithm: Robustness and Convergence

Mingyi Hong, Alfredo Garcia

## Abstract

The convergence properties of the Iterative water-filling (IWF) based algorithms ([1], [2], [3]) have been derived in the ideal situation where the transmitters in the network are able to obtain the *exact* value of the interference plus noise (IPN) experienced at the corresponding receivers *in each iteration* of the algorithm. However, these algorithms are not robust because they diverge when there is *time-varying* estimation error of the IPN, a situation that arises in real communication system. In this correspondence, we propose an algorithm that possesses convergence guarantees in the presence of various forms of such time-varying error. Moreover, we also show by simulation that in scenarios where the interference is strong, the conventional IWF diverges while our proposed algorithm still converges.

## I. INTRODUCTION

### A. The IWF Algorithm

The Iterative Water-Filling algorithm has been first proposed by Yu et al in [1] to solve the power allocation problem in DSL network, and it has since been applied to various areas in communications and signal processing to obtain solutions for network power allocation problems (see, e.g. [3], [4], [5], [6] and the references therein).

We consider an application of the IWF algorithm to the resource allocation problem in wireless communication network, where there are  $N$  users and  $K$  subchannels; each user is a transmitter-receiver pair that tries to communicate with each other. Define the sets  $\mathcal{N} \triangleq \{1, \dots, N\}$ , and  $\mathcal{K} \triangleq \{1, \dots, K\}$ ; let  $\{S_i\}_{i \in \mathcal{N}}$  denote the set of users in the network; let  $p_i(k)$  denote the amount of power  $S_i$  transmits on channel  $k$ ; let  $\mathbf{p}_i \triangleq [p_i(1), \dots, p_i(K)]^\top$ ,  $\mathbf{p}_{-i} \triangleq [\mathbf{p}_1^\top, \dots, \mathbf{p}_{i-1}^\top, \mathbf{p}_{i+1}^\top, \dots, \mathbf{p}_N^\top]^\top$  and  $\mathbf{p} \triangleq [\mathbf{p}_1^\top, \dots, \mathbf{p}_N^\top]^\top$ . The channel gain between the transmitter of  $S_i$  to the receiver of  $S_j$  on channel  $k$  is denoted by  $|H_{i,j}(k)|^2$ .

M. Hong and A. Garcia are with the Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, 22903 USA (e-mail: mh4tk@virginia.edu, ag7s@virginia.edu)

The power of the environmental noise experienced at  $S_i$ 's receiver on channel  $k$  is denoted by  $n_i(k)$ . We assume that there is no interference cancelation performed at the receivers, and the interference caused by the other users is considered as noise. Then the signal to interference plus noise ratio (SINR) measured at the receiver of  $S_i$  on channel  $k$  can be expressed as:  $SINR_i(k) = \frac{|H_{i,i}(k)|^2 p_i(k)}{n_i(k) + \sum_{j \neq i} |H_{j,i}(k)|^2 p_j(k)}$ .

Using Shannon's capacity, the maximum transmission rate achievable for  $S_i$  can be expressed as:  $R_i(\mathbf{p}_i, \mathbf{p}_{-i}) = \sum_{i=1}^K \log(1 + SINR_i(k))$ . We consider the following constraints for each user: **[C-1]** each  $S_i$  has limited power budget, i.e.,  $0 \leq \sum_{k=1}^K p_i(k) \leq \bar{p}_i, \forall i \in \mathcal{N}$ ; **[C-2]** we require  $0 \leq p_i(k) \leq p_{mask}(k), \forall k \in \mathcal{K}$  and  $i \in \mathcal{N}$ . As such, we use  $\mathcal{P}_i$  to denote the set of feasible power allocations for  $S_i$ :  $\mathcal{P}_i \triangleq \left\{ \mathbf{p}_i : \sum_{k=1}^K p_i(k) \leq \bar{p}_i, 0 \leq p_i(k) \leq p_{mask}(k), \forall k \in \mathcal{K} \right\}$ .

Dynamic power allocation in this network can be formulated as a non-cooperative game where each user  $S_i$  is interested in maximizing its own rate when deciding how to allocate its power across the spectrum, i.e.,  $S_i$  wants to find  $\mathbf{p}_i^* \in \mathcal{P}_i$  such that  $\mathbf{p}_i^* \in \arg \max_{\mathbf{p}_i \in \mathcal{P}_i} R_i(\mathbf{p}_i, \mathbf{p}_{-i})$ . A Nash Equilibrium (NE) can be expressed as the set of power profiles  $\{\mathbf{p}_i^*\}_{i \in \mathcal{N}}$  satisfying the set of conditions:  $\mathbf{p}_i^* \in \arg \max_{\mathbf{p}_i \in \mathcal{P}_i} R_i(\mathbf{p}_i, \mathbf{p}_{-i}^*) \forall i \in \mathcal{N}$ . The IWF and its various extensions are essentially policies for the players to jointly reach a NE of this game in a distributed manner.

In the IWF, the transmitters iteratively adjust their transmission power levels to maximize their own transmission rate. Specifically, in iteration  $t + 1$ , each user  $S_i$  computes  $\{p_i^{t+1}(k)\}_{k \in \mathcal{K}}$  as follows:

$$\begin{aligned} p_i^{t+1}(k) &= \arg \max_{\mathbf{p}_i \in \mathcal{P}_i} R_i(\mathbf{p}_i, \mathbf{p}_{-i}^t) \\ &= \left[ \sigma_i - (\bar{n}_i(k) + \sum_{j \neq i} |\bar{H}_{j,i}(k)|^2 p_j^t(k)) \right]_0^{p_{mask}(k)} \\ &= [\sigma_i - IPN_i^t(k)]_0^{p_{mask}(k)} \triangleq \Phi_i^k(\mathbf{p}_{-i}^t) \end{aligned} \quad (1)$$

where  $\sigma_i$  is the dual variable associated with the total power constraint for user  $i$ , and it is also referred to as the "water level" in the traditional water-filling algorithm;  $|\bar{H}_{j,i}(k)|^2$  and  $\bar{n}_i(k)$  are defined as  $|\bar{H}_{j,i}(k)|^2 \triangleq \frac{|H_{j,i}(k)|^2}{|H_{i,i}(k)|^2}$  and  $\bar{n}_i(k) \triangleq \frac{n_i(k)}{|H_{i,i}(k)|^2}$ , respectively;  $IPN_i^t(k)$  is defined as the normalized total interference plus noise (IPN) for user  $S_i$  on channel  $k$  at time  $t$ :  $IPN_i^t(k) \triangleq \bar{n}_i(k) + \sum_{j \neq i} |\bar{H}_{j,i}(k)|^2 p_j^t(k)$ . This quantity is measured at the receivers and fed back to their corresponding transmitters in each iteration  $t$  before  $p_i^{t+1}(k)$  is computed. Define  $\Phi_i(\mathbf{p}_{-i}) \triangleq [\Phi_i^1(\mathbf{p}_{-i}), \dots, \Phi_i^K(\mathbf{p}_{-i})]^\top$ , and let  $\Phi(\mathbf{p}) \triangleq [\Phi_1(\mathbf{p}_{-1}), \dots, \Phi_N(\mathbf{p}_{-N})]^\top$ . The function  $\Phi(\cdot)$  is called the *water-filling operator* of the system, and the IWF algorithm can be written concisely:  $\mathbf{p}^{t+1} = \Phi(\mathbf{p}^t)$ . If the algorithm reaches a power profile  $\mathbf{p}^*$  such that  $\mathbf{p}^* = \Phi(\mathbf{p}^*)$ , we say the IWF converges.

Sufficient conditions for convergence of the IWF algorithm and its various extensions have been widely

studied, for example, in [4], [6], [7]. Essentially, if the interference received (generated) at the receiver (transmitter) of each user is weak enough compared with the desired signal, then the IWF converges. When these conditions are not met, it is possible that the IWF diverges [8].

### B. The Uncertainty of IPN and the Water-Filling Operator

One of the key assumptions of the IWF based algorithms is that the receivers can *always* get the exact values of the IPN on each channel in each iteration of the algorithm, and fed back to the transmitters. This assumption is not valid in real communication systems because the power of the noise/interference experienced at the receivers needs to be *estimated* in each iteration, thus is subject to time-varying estimation errors [9], [10]. Therefore, in each iteration of the algorithm, we can only obtain a noisy version of the true solution of (1), referred to as the *noisy water-filling solution*, as:

$$p_i^{t+1}(k) = \left[ \hat{\sigma}_i - \widehat{IPN}_i^t(k) \right]_0^{p_{mask}(k)} \triangleq \widehat{\Phi}_i^k(\mathbf{p}_{-i}^t) \quad (2)$$

where  $\widehat{IPN}_i^t(k)$  is the noisy (estimated) IPN for user  $S_i$  on channel  $k$ . Note that the uncertainty of the IPN leads to the inaccuracy of the dual variable, as now it should satisfy  $\sum_{k=1}^K \left[ \hat{\sigma}_i - \widehat{IPN}_i^t(k) \right]_0^{p_{mask}(k)} = \bar{p}_i$ , and  $\sigma_i \neq \hat{\sigma}_i$  in general.

There is little work in the literature that addresses the impact of such time-varying uncertainty of the IPN on the performance of the IWF algorithm. In [3], [4], a “relaxed” version of IWF (R-IWF) was proposed to heuristically deal with inaccurate IPN levels. In each iteration, the transmission power levels are computed as  $\mathbf{p}^{t+1} = (1 - \lambda)\mathbf{p}^t + \lambda\Phi(\mathbf{p}^t)$ , where the  $\lambda \in (0, 1]$  is a free parameter. Although it has been shown in [4] that this algorithm converges under similar conditions as the IWF in situations *without* IPN uncertainty, the effect of this algorithm in the presence of IPN inaccuracy is not clear, and as we will see later in the simulation section, the performance of R-IWF depends strongly on the choice of  $\lambda$ . In [11], a robust version of IWF is proposed to deal with errors related to changes in the number of users and their mobility. The algorithm guarantees an acceptable level of performance under worst case conditions (i.e., the maximum possible error of the IPN). This algorithm trades performance in favor of robustness, thus the equilibrium solution obtained is generally less efficient than that of the original IWF. In our work, we are concerned with reaching the equilibrium solution of the original IWF in the presence of IPN uncertainty. In [12], the authors provide a probabilistically robust IWF to deal with the quantization errors of the IPN at the receiver of each user. In this algorithm, users allocate their powers to maximize their total rate for a large fraction of the error realization. However, a specific distribution of the error process is assumed in the derivation of the algorithm, and such statistical information is usually

not available in practice (as suggested in section V of [11]). A recent work [13] proposes algorithms for system with finite-state Markov channel in interference network. The channel itself is modeled as time-varying in this work, and the objective is to track the time-varying equilibria. In the present paper, uncertainty of the channel is due to imperfect receiver estimation of the value of IPN as opposed to changes in the state of the channel.

In this correspondence, we propose an extension of the IWF algorithm that is robust in the presence of time-varying IPN uncertainties. Specifically, we model the uncertainty regarding to the IPN as time-varying added noises, and show that the proposed algorithm converges with probability 1 under some conditions on the channel gains and the noise process. We verify the above claim by simulation, and demonstrate the advantage of the proposed algorithm with respect to the original IWF and the R-IWF. Additionally, we show by simulation that in some strong interference channels where the conventional IWF algorithm diverges, our proposed algorithm still converges. This last result indicates that the convergence condition of our algorithm may be further relaxed.

This correspondence is organized as follows. In section II we introduce the proposed algorithm and provide convergence analysis. In section III we demonstrate the performance of the proposed algorithm and compare the results with conventional IWF. This correspondence concludes in section IV.

## II. PROPOSED ALGORITHM AND CONVERGENCE RESULTS

In the proposed algorithm, in each iteration  $t$ , all the users compute their power allocations as follows:

- 1) Obtain  $\{\widehat{IPN}_i^t(k)\}_{k \in \mathcal{K}}$ , and calculate the noisy water-filling solution  $\widehat{\Phi}_i(\mathbf{p}_{-i}^t)$ .
- 2) Calculate the power output according to the following policy:

$$\mathbf{p}_i^{t+1} = \begin{cases} \widehat{\Phi}_i(\mathbf{p}_{-i}^t) & \text{for } t = 0 \\ (1 - \alpha_t)\mathbf{p}_i^t + \alpha_t\widehat{\Phi}_i(\mathbf{p}_{-i}^t) & \text{for } t \geq 1 \end{cases} \quad (3)$$

where the elements of  $\widehat{\Phi}_i(\mathbf{p}_{-i}^t)$  are defined in (2). The sequence  $\{\alpha_t : 0 < \alpha_t \leq 1\}_{t=0}^{\infty}$  satisfies the following (define  $\alpha_0 = 1$ ):

$$\lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t = \infty, \quad \lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t^2 < \infty. \quad (4)$$

Note that from the last inequality in (4), we have  $\lim_{t \rightarrow \infty} \alpha_t = 0$ . The update procedure (3) is essentially Mann's iterations (see [14] for its properties), which is designed for situations where conventional iterative methods for finding the fixed point of a self-mapping (say Picard's method) fail. If we choose  $\alpha_t = \frac{1}{t+1}$ , then the update policy in (3) can be rewritten as:  $\mathbf{p}_i^{T+1} = \frac{1}{T+1} \sum_{t=0}^T \widehat{\Phi}_i(\mathbf{p}_{-i}^t)$ . Clearly  $\mathbf{p}_i^{T+1}$  is an *average* of the history of  $S_i$ 's water-filling solution, hence the name of Average Iterative Water-Filling (A-IWF)

for the proposed algorithm. This algorithm maintains the distributed nature of the original IWF, because in each iteration  $t + 1$ ,  $S_i$  only needs to know the set of IPN  $\{IPN_i^t(k)\}_{k \in \mathcal{K}}$  as well as its own power allocation  $\{p_i^t(k)\}_{k \in \mathcal{K}}$  in iteration  $t$  (both of which can be obtained locally by  $S_i$ ), but *does not* need to know the transmission power profiles of other users.

We see that the main difference between the proposed algorithm and the previously mentioned R-IWF is that we use a set of *diminishing* and *iteration dependent* stepsize  $\{\alpha_t\}_{t=0}^\infty$  that satisfies (4), instead of the fixed stepsize  $\lambda$ . We will see later that it is exactly these properties of the  $\{\alpha_t\}_{t=0}^\infty$  that guarantee the convergence of A-IWF under IPN uncertainty.

We model the noisy IPN for user  $S_i$  on channel  $k$  as:  $\widehat{IPN}_i^t(k) = IPN_i^t(k) + \epsilon_i^t(k)$ , where  $\epsilon_i^t(k)$  represents the estimation error of the true value  $IPN_i^t(k)$ . Let  $\epsilon_i \triangleq [\epsilon_i(1), \dots, \epsilon_i(K)]^\top$ , and  $\epsilon \triangleq [\epsilon_1, \dots, \epsilon_N]^\top$ . Let  $\mathcal{F}_i^T$  be defined as the filtration generated by  $\mathbf{p}_i^{T+1} \cup \{\mathbf{p}_i^t, \widehat{\Phi}_i(\mathbf{p}_{-i}^t)\}_{t=0}^T$ . We assume the error process to be zero mean, i.e.,  $E[\epsilon_i^t(k) | \mathcal{F}_i^{t-1}] = 0$ . This assumption is reasonable because conditioning on the knowledge of the desired signal ( $\mathbf{p}_i^t$  in our case), the estimation error of  $IPN_i^t(k)$ ,  $\epsilon_i^t(k)$  can indeed be viewed as a zero mean random variable using most conventional estimators (see Section V of [9] for detailed comparison of estimation biases for different algorithms). The above model is very general in the sense that we do not assume the explicit forms of the algorithms that perform the estimation, nor do we require that the error process  $\{\epsilon_i^t(k)\}_{t=1}^T$  be independent with the history of IPN up to time  $T$ , i.e., our model allows  $\widehat{IPN}_i^t(k)$  to be calculated based on the previous or the current observations made by the receiver of  $S_i$ .

In the following, we use “w. p. 1” to abbreviate “with probability 1”. We need the following definition before introducing Lemma 1, which characterizes the noisy version of the water-filling operator  $\widehat{\Phi}(\mathbf{p})$ . For any positive  $N \times 1$  vector  $\mathbf{w} \triangleq [w_1, \dots, w_N]^\top$  and the operator  $\widehat{\Phi}(\mathbf{p})$ , the (vector) *block-maximum norm*  $\|\cdot\|_{2,block}^{\mathbf{w}}$  is defined as [15]:  $\|\widehat{\Phi}(\mathbf{p})\|_{2,block}^{\mathbf{w}} \triangleq \max_{i \in \mathcal{N}} \frac{\|\widehat{\Phi}_i(\mathbf{p}_{-i})\|_2}{w_i}$ .

*Lemma 1: Define a  $N \times N$  matrix  $\Upsilon$  related to the channel gains as:*

$$[\Upsilon]_{i,j} \triangleq \begin{cases} \max_{k \in \mathcal{K}} |\bar{H}_{j,i}(k)|^2 & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}. \quad (5)$$

*Let  $\rho(\Upsilon)$  be the spectral radius of the matrix  $\Upsilon$ . Then if  $\rho(\Upsilon) < 1$ , there must exist a positive vector  $\bar{\mathbf{w}}$ , and a constant  $\beta$  that satisfies  $0 < \beta < 1$ , such that for any feasible  $\mathbf{p}^1, \mathbf{p}^2 \in \mathcal{P}$ ,*

$$\|\widehat{\Phi}(\mathbf{p}^1) - \widehat{\Phi}(\mathbf{p}^2)\|_{2,block}^{\bar{\mathbf{w}}} \leq \beta \|\mathbf{p}^1 - \mathbf{p}^2\|_{2,block}^{\bar{\mathbf{w}}} + \|\epsilon\|_{2,block}^{\bar{\mathbf{w}}}. \quad (6)$$

*Proof:* The Proof is similar to Proposition 2 of [4]. Please see Appendix A for detail. ■

We note here that it has been proven by [4], that when  $\rho(\Upsilon) < 1$  is true, the original water-filling operator  $\Phi(\mathbf{p})$  is a contraction with coefficient  $\beta < 1$ , and hence has a *unique* fixed point, i.e., there exists a unique  $\mathbf{p}^* \in \mathcal{P}$  such that  $\mathbf{p}^* = \Phi(\mathbf{p}^*)$ .

We then characterize the convergence property of the A-IWF algorithm under two different assumptions of the noise process  $\{\epsilon^t\}_{t=0}^\infty$ . For simplicity of notation, in the following, we use  $\|\cdot\|$  to denote the norm  $\|\cdot\|_{2,block}^{\bar{\mathbf{w}}}$ , where  $\bar{\mathbf{w}}$  is the positive vector obtained from the proof of Lemma 1.

*Theorem 1:* Assume  $\rho(\Upsilon) < 1$ ,  $\{\alpha_t\}_{t=0}^\infty$  satisfies (4), and  $\{\epsilon^t\}_{t=0}^\infty$  satisfies  $\sum_{t=1}^\infty \alpha_t \|\epsilon^t\| < \infty$ , w. p. 1. Then the sequence of power profiles  $\{\mathbf{p}^t\}_{t=1}^\infty$  generated by the A-IWF algorithm converges to the unique fixed point of the original mapping  $\Phi(\cdot)$ , denoted by  $\mathbf{p}^*$ . More precisely, we have:  $\|\mathbf{p}^t - \mathbf{p}^*\| \rightarrow 0$  w. p. 1.

*Proof:* Please see Appendix B for proof. ■

*Theorem 2:* Assume  $\rho(\Upsilon) < 1$ , and  $\{\alpha_t\}_{t=0}^\infty$  satisfies (4), and the error process satisfies  $\lim_{t \rightarrow \infty} \|\epsilon^t\| = 0$ , w. p. 1. Then we have:  $\|\mathbf{p}^t - \mathbf{p}^*\| \rightarrow 0$  w. p. 1.

*Proof:* Please see Appendix C for proof. ■

At this point, we would like to give some remarks regarding to the above convergence results.

*Remark 1:* The condition  $\rho(\Upsilon) < 1$ , which is a restriction on the channel gains, coincides with the condition that ensures the convergence of IWF without the IPN uncertainties in Theorem 1 of [4]. We refer the readers to [4] for physical interpretation as well as the comparison of this condition with other similar conditions derived in the literature, e.g., those in [6] and [7].

*Remark 2:* We will show in section III-B that in many cases when  $\rho(\Upsilon) < 1$  is not satisfied, our algorithm still converges. This suggests that the A-IWF algorithm may need more relaxed convergence conditions than the one stated in this correspondence. We will leave this task as a future research topic.

Theorem 1 and Theorem 2 differ in their respective restrictions on the error process  $\{\epsilon^t\}_{t=0}^\infty$ , as technically the conditions  $\sum_{t=0}^\infty \alpha_t \|\epsilon^t\| < \infty$  and  $\lim_{t \rightarrow \infty} \|\epsilon^t\| = 0$  do not imply each other. Although these conditions require that the error process be *diminishing*, we do observe in our simulations (to be shown in Section III) that the A-IWF converges in the presence of more general forms of noises, for example noises with zero mean and bounded second moment. This observation leads us to believe that the above conditions on the error process are overly restrictive. Such belief is partially justified as follows.

Assume  $E[\epsilon_i^t | \mathcal{F}_i^{t-1}] = 0$ , and  $\epsilon_i^t$  has bounded second moment for all  $i$ . Further assume  $\hat{\Phi}(\mathbf{p}^t)$  can be approximated as:  $\hat{\Phi}(\mathbf{p}^t) = \Phi(\mathbf{p}^t) + \boldsymbol{\xi}^t$ , where the elements of the *bias vector*  $\boldsymbol{\xi}^t$  satisfies:

$$E[\xi_i^t(k) | \mathcal{F}_i^{t-1}] = 0 \text{ and } E[(\xi_i^t(k))^2 | \mathcal{F}_i^{t-1}] < \infty. \quad (7)$$

Then we have the following convergence result. See Appendix D for the proof.

*Theorem 3:* Suppose  $\widehat{\Phi}(\mathbf{p}^t)$  is approximated as  $\widehat{\Phi}(\mathbf{p}^t) = \Phi(\mathbf{p}^t) + \boldsymbol{\xi}^t$  with the elements of  $\boldsymbol{\xi}^t$  satisfies (7). If  $\Phi(\cdot)$  is a contraction with constant  $\beta$ , and if  $\{\alpha_t\}_{t=1}^\infty$  satisfy (4), then we have:  $\|\mathbf{p}^t - \mathbf{p}^*\| \rightarrow 0$  w. p. 1.

Theorem 3 essentially says that if the above approximation of the noisy water-filling solution is accurate, then we only require the error process  $\{\boldsymbol{\epsilon}^t\}_{t=0}^\infty$  to have mean zero and bounded second moments to ensure the convergence of the algorithm. Note that in this case the bias vector  $\boldsymbol{\xi}^t$  summarizes the uncertainties regarding both the IPNs and the dual variables. The key assumption here is that  $E[\xi_i^T(k)|\mathcal{F}_i^{T-1}] = 0 \forall i, k$ , i.e., based on all the knowledge it has for the evolution of the algorithm until time  $T - 1$ , a particular user  $S_i$  predicts that the biases  $\{\xi_i^T(k)\}_k$  are zero mean. The following empirical experiments show that such assumption is approximately true.

Consider a network with 10 users and 32 channels. Let  $\bar{p}_i = 10$ ,  $p_{mask}(k) = 3$ ,  $\forall k \in \mathcal{K}$ . We define the bias of the noisy water-filling solution as:

$$\begin{aligned} \xi_i(k) &\triangleq \widehat{\Phi}_i^k(\mathbf{p}_{-i}) - \Phi_i^k(\mathbf{p}_{-i}) \\ &= [\widehat{\sigma}_i - IPN_i(k) - \epsilon_i(k)]_0^{p_{mask}(k)} - [\sigma_i - IPN_i(k)]_0^{p_{mask}(k)}. \end{aligned}$$

We simplify the analysis a bit by assuming the bias process to be Markovian, i.e.,  $E[\xi_i^T(k)|\mathcal{F}_i^{T-1}] = E[\xi_i^T(k)|\mathbf{p}_i^T]$ . We investigate the distribution of  $\{E[\xi_i(k)|\mathbf{p}_i]\}_{i,k}$ . Define the variance of noise  $\epsilon_i(k)$  as  $var_i(k)$ ; introduce a term called Interference Error Ratio (IER) to quantify the strength of the IPN error  $\epsilon$ :  $IER_i(k) \triangleq 10 \log_{10} \left( \frac{IPN_i(k)}{var_i(k)} \right)$ . We fix  $IER = 10\text{dB}$  during the experiment. As  $E[\xi_i(k)|\mathbf{p}_i]$  is a function of  $\mathbf{p}_i$ , we fix  $\{\mathbf{p}_i \in \mathcal{P}_i\}_{i \in \mathcal{N}}$ , and obtain an estimate of  $\{E[\xi_i(k)|\mathbf{p}_i]\}_{i,k}$ , denoted by  $\{M_i(k)\}_{i,k}$ , by doing the follows: 1) generate the channel gains  $\{|H_{i,j}(k)|^2\}$  randomly; 2) generate  $L$  samples of IPN noise vectors  $\{\boldsymbol{\epsilon}^l\}_{l=1}^L$  by:  $\epsilon_i^l(k) \sim N(0, var_i(k))$ ,  $\forall i, k, y$ ; 3) obtain the bias  $\{\boldsymbol{\xi}^l\}_{l=1}^L$  according to its definition above; 4) calculate  $M_i(k) = \frac{1}{L} \sum_{l=1}^L \xi_i^l(k)$ ,  $\forall i, k$ . We repeat the above procedure for 1,000 times with randomly generated sets of  $\{\mathbf{p}_i \in \mathcal{P}_i\}_{i \in \mathcal{N}}$ , and plot the empirical distribution of  $\{E[\xi_i(k)|\mathbf{p}_i]\}_{i,k}$  in Fig. 1 (different graphs in Fig. 1 represent the results obtained by experiments using different  $L$ ). We see that when the estimates  $\{M_i(k)\}_{i,k}$  are getting more accurate with larger number of samples (larger  $L$ ), the empirical distribution of  $\{E[\xi_i(k)|\mathbf{p}_i]\}_{i,k}$  is getting more concentrated at zero. Thus we conjecture that asymptotically with  $L \rightarrow \infty$ ,  $E[\xi_i(k)|\mathbf{p}_i]$  can be approximated as zero for all  $i$  and  $k$ .

*Remark 3:* We give some remarks comparing the convergence conditions of conventional IWF and A-IWF under uncertainty. From [16] (Chapter 12, Th. 12.2.1–12.2.5) we see that the condition  $\lim_{t \rightarrow \infty} \|\boldsymbol{\epsilon}^t\| = 0$  in Theorem 2 is sufficient and necessary for the conventional IWF to converge to the fixed point without performing averaging. However, the conventional IWF diverges under condition  $\sum_{t=1}^\infty \alpha_t \|\boldsymbol{\epsilon}^t\| < \infty$  in Theorem 1, because this condition is not equivalent to  $\lim_{t \rightarrow \infty} \|\boldsymbol{\epsilon}^t\| = 0$ . Moreover, from Th. 12.2.5 in



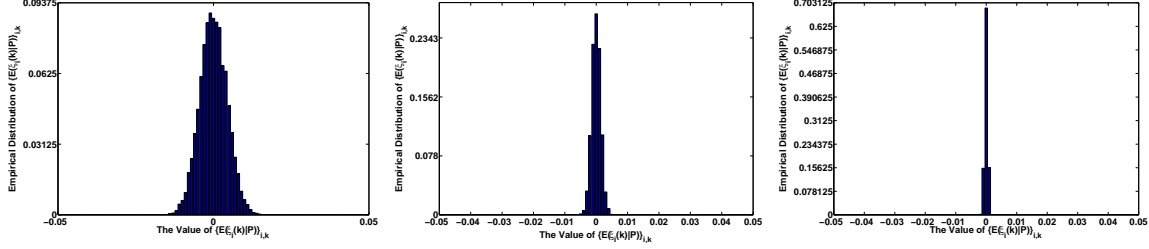


Fig. 1. Empirical distribution of  $\{E[\xi_i(k)|\mathbf{p}]\}_{i,k}$ . Left:  $L = 1,000$ ; Middle:  $L = 10,000$ ; Right:  $L = 100,000$ .

[16], under the assumption in Theorem 3, the conventional IWF produces a sequence that finally stays in a ball around the fixed point. However, the radius of such ball is increasing with  $\frac{\max_t \|\xi^t\|}{1-\beta}$ . Notice that in this case  $\|\xi^t\|$  needs not to be decreasing, thus the maximum possible error of the conventional IWF may be large (consider when  $\beta$  is close to 1).

### III. SIMULATION RESULTS

In this section we conduct three experiments to demonstrate the properties of the A-IWF algorithm.

#### A. Performance with Estimation Error

We simulate a network with 10 randomly located users, and 64 channels. We choose the noise to be a zero mean Gaussian random variable as  $\epsilon_i^t(k) \sim N(0, \text{var}_i^t(k))$ ; we choose the  $IER$  for all the users on all the channels to be  $IER_i^t(k) = 20\text{dB}$ ,  $15\text{dB}$ ; we choose the channel gains  $\{|H_{i,j}(k)|^2\}$  randomly and appropriately such that  $\rho(\Upsilon) < 1$  is satisfied; we choose  $\alpha_t = \frac{1}{t+1}$ . For ease of demonstration, different algorithms are examined with the same starting points.

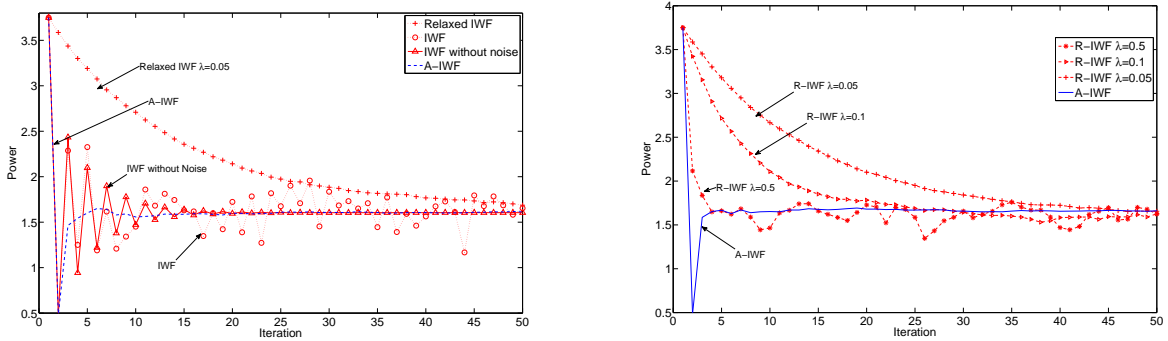


Fig. 2. Comparison of the output for different algorithms,  $IER=20\text{dB}$ .

In Fig. 2, we show the power output produced by various algorithms of a particular user on a particular channel, with  $IER = 20\text{dB}$ . It is clear that in the presence of estimation error, the IWF algorithm



produces a sequence of noisy power profiles which exhibits no sign of convergence. We also show the performance of IWF algorithm *without* estimation error, for the purpose of comparison. It is seen that the A-IWF algorithm converges to the unique NE predicted by the IWF (without estimation error) quickly. In Fig. 2, we also show the output of the R-IWF algorithm with various values of  $\lambda$ . We observe that when  $\lambda$  is large, the output is still noisy, while when  $\lambda$  is small, the convergence is slow. The point is that the choice of  $\lambda$  is important for the performance of R-IWF, but it is difficult to correctly choose  $\lambda$  to guarantee both robustness and fast convergence. In Fig. 3, we compare the selected power profiles of R-IWF and A-IWF when  $IER = 15dB$ .

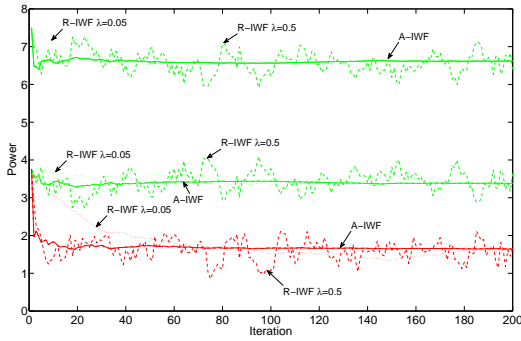


Fig. 3. Comparison of different algorithms, IER=15dB.

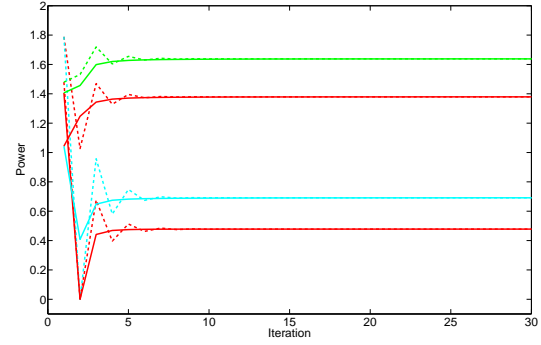


Fig. 4. Comparison of convergence speed of IWF and A-IWF.

### B. Performance with Strong Interference

As stated above, the convergence of the IWF in ideal situations usually depends on the weak interference condition. It is observed that in the system with strong interference, IWF algorithm diverges [8]. In the following simulation, we demonstrate several scenarios in which the IWF diverges, but the A-IWF algorithm converges. The purpose of these simulations is to argue that the A-IWF may need weaker conditions for convergence.

Consider the following scenario of strong interference (example 5 in [8]). Suppose there are 3 users and 2 channels in the system, with channel matrices  $\mathbf{H}(k)$  expressed as follows:

$$\mathbf{H}(1) = \mathbf{H}(2) = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$$

where each element of the matrix  $\mathbf{H}(k)$  is defined as  $[\mathbf{H}(k)]_{i,j} \triangleq |H_{i,j}(k)|^2$ . Set the noise power on channel 1 to be  $\sigma^2$ , the noise on channel 2 set to be  $\sigma^2 + \bar{p}_i$ , with  $\bar{p}_i = 10$ , for all  $i \in \mathcal{N}$ . There is a

unique NE of this game, in which each user allocates two-thirds of its power to channel 1 and the rest to channel 2. The left hand side part of Fig. 5 shows the power profiles of user 1 on channel 1 that are produced by different algorithms (with the same starting point). It is seen that the IWF algorithm oscillates, while the A-IWF algorithm converges quickly. Similar results are obtained in the right hand side part of Fig. 5 with the following settings:

$$\mathbf{H}(1) = \begin{pmatrix} 1 & 2 & 4 \\ 4 & 1 & 2 \\ 2 & 4 & 1 \end{pmatrix}, \mathbf{H}(2) = \begin{pmatrix} 2 & 3 & 5 \\ 3 & 2 & 5 \\ 5 & 3 & 2 \end{pmatrix} \quad (8)$$

and the noise power on both channels set to be the same. We observe again that the performance of R-IWF algorithm is very sensitive to the choice of  $\lambda$ : when  $0.6 \leq \lambda \leq 1$ , the output oscillates; when  $0 \leq \lambda \leq 0.5$ , the output converges, with larger  $\lambda$  for faster convergence. However, it is not clear what rules one should follow in general to select such critical parameter.

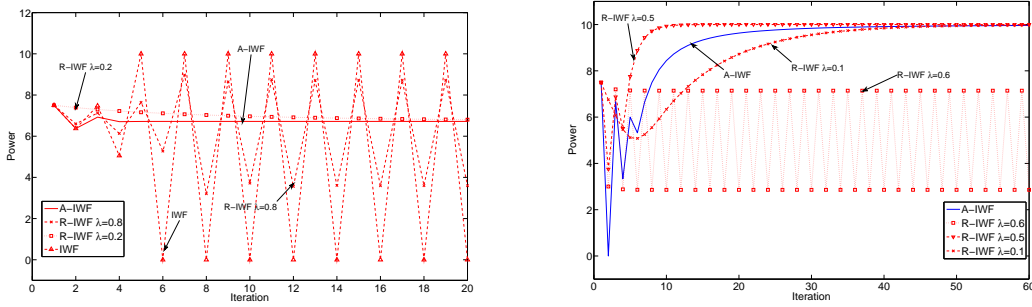


Fig. 5. Convergence properties of difference algorithms in strong interference channels.

### C. Convergence In Ideal Cases

Questions may arise as to how does the A-IWF perform in situations when the water-filling solution in (1) can be carried out accurately. As shown in [4], the IWF algorithm converges *linearly* in this ideal scenario. Theoretically, we can only show that A-IWF converges *sublinearly* in ideal scenario, i.e.,  $\lim_{t \rightarrow \infty} \frac{\|\mathbf{p}^{t+1} - \mathbf{p}^*\|}{\|\mathbf{p}^t - \mathbf{p}^*\|} = 1$ . However, we observe in various randomly generated channel gains and random starting points of the algorithms, that the A-IWF algorithm seems to always converge as fast as the IWF algorithm. Fig. 4 shows such an instance of this experiment. In this figure, we compare the power output of selected users on selected channels (in a network with 10 users and 64 channels) generated by the IWF and the A-IWF. It takes less than 10 iterations before two algorithms agree with each other. Note that the dotted lines represent the output of the IWF algorithms and the solid lines represent the output of the A-IWF algorithm.

#### IV. CONCLUSION

In this correspondence, we proposed an extension to the IWF algorithm which is more robust and has better convergence properties. We proved that the proposed algorithm converges w. p. 1 under suitable assumptions. We argue that this algorithm is indeed robust against time-varying estimation error of the power of interference plus noise that is needed for the computation of the IWF computation. We also show by simulation that the proposed algorithm converges when strong interferences are present in the communication channel, a scenario in which the IWF algorithm diverges. An interesting future research topic is to develop a possibly more general condition for the convergence of the proposed algorithm.

#### APPENDIX A

##### PROOF OF LEMMA 1

*Proof:* Define  $\bar{\mathbf{H}}_{j,i} = \text{diag}(|\bar{H}_{j,i}(1)|^2, \dots, |\bar{H}_{j,i}(K)|^2)$ ; define  $\mathbf{IPN}_i = [IPN_i(1), \dots, IPN_i(K)]^T$ , and define  $\widehat{\mathbf{IPN}}_i$  similarly. From Corollary 3 of [4], we have that the water-filling operator  $\Phi_i(\mathbf{p}_{-i})$  can be expressed as the projection of  $-\mathbf{IPN}_i$  onto the space  $\mathcal{P}_i$ , i.e.,  $\Phi_i(\mathbf{p}_{-i}) = [-\mathbf{IPN}_i]_{\mathcal{P}_i}$ . Similarly, we have that  $\widehat{\Phi}_i(\mathbf{p}_{-i}) = [-\widehat{\mathbf{IPN}}_i]_{\mathcal{P}_i}$ . Consequently, we have:

$$\begin{aligned} \|\widehat{\Phi}_i(\mathbf{p}_{-i}^1) - \Phi_i(\mathbf{p}_{-i}^2)\|_2 &\stackrel{(a)}{\leq} \left\| -\sum_{j \neq i} \bar{\mathbf{H}}_{j,i} \mathbf{p}_j^1 - \boldsymbol{\epsilon}_i + \sum_{j \neq i} \bar{\mathbf{H}}_{j,i} \mathbf{p}_j^2 \right\|_2 \\ &\leq \sum_{j \neq i} \|\bar{\mathbf{H}}_{j,i}\|_2 \|\mathbf{p}_j^1 - \mathbf{p}_j^2\|_2 + \|\boldsymbol{\epsilon}_i\|_2 \\ &\stackrel{(b)}{=} \sum_{j \neq i} \left( \max_k |\bar{H}_{j,i}(k)|^2 \right) \|\mathbf{p}_j^1 - \mathbf{p}_j^2\|_2 + \|\boldsymbol{\epsilon}_i\|_2, \quad \forall i \in \mathcal{N} \end{aligned} \quad (9)$$

where (a) is because of the non-expansiveness of the projection operator under Euclidean norm; (b) is due to the fact that the 2-norm of a diagonal matrix equals to the maximum absolute value of its diagonal elements. Define  $e_{\Phi_i} \triangleq \|\widehat{\Phi}_i(\mathbf{p}_{-i}^1) - \Phi_i(\mathbf{p}_{-i}^2)\|_2$ ,  $e_i \triangleq \|\mathbf{p}_i^1 - \mathbf{p}_i^2\|_2$ , and let  $\mathbf{e}_{\Phi} \triangleq [e_{\Phi_1}, \dots, e_{\Phi_N}]^T$ ,  $\mathbf{e} \triangleq [e_1, \dots, e_N]^T$ , and  $\mathbf{e}_{\epsilon} \triangleq [\|\boldsymbol{\epsilon}_1\|_2, \dots, \|\boldsymbol{\epsilon}_N\|_2]^T$ .

In order to proceed, we define the vector weighted maximum norm [15] as:

$$\|\mathbf{x}\|_{\infty, \text{vec}}^{\mathbf{w}} \triangleq \max_i \frac{|x_i|}{w_i}, \quad \mathbf{w} > \mathbf{0}, \mathbf{x} \in \mathbb{R}^N \quad (10)$$

and the matrix weighted maximum norm as:

$$\|\mathbf{A}\|_{\infty, \text{mat}}^{\mathbf{w}} \triangleq \max_i \frac{1}{w_i} \sum_{j=1}^N |[\mathbf{A}]_{i,j}| w_j, \quad \mathbf{w} > \mathbf{0}, \mathbf{A} \in \mathbb{R}^{N \times N}. \quad (11)$$

Notice, that from the definition of norm  $\|\cdot\|_{\infty,vec}^{\mathbf{w}}$ ,  $\|\cdot\|_{\infty,mat}^{\mathbf{w}}$  and the block-maximum norm, we have the following equivalence:

$$\begin{aligned} \|\mathbf{e}\|_{\infty,vec}^{\mathbf{w}} &= \max_i \frac{|e_i|}{w_i} = \max_i \frac{\|\mathbf{p}_i^1 - \mathbf{p}_i^2\|}{w_i} = \|\mathbf{p}^1 - \mathbf{p}^2\|_{2,block}^{\mathbf{w}} \\ \|\mathbf{e}_{\Phi}\|_{\infty,vec}^{\mathbf{w}} &= \max_i \frac{|e_{\Phi_i}|}{w_i} = \|\hat{\Phi}(\mathbf{p}^1) - \Phi(\mathbf{p}^2)\|_{2,block}^{\mathbf{w}} \\ \|\mathbf{e}_{\epsilon}\|_{\infty,vec}^{\mathbf{w}} &= \max_i \frac{\|\epsilon_i\|_2}{w_i} = \|\epsilon\|_{2,block}^{\mathbf{w}}. \end{aligned} \quad (12)$$

The set of  $N$  inequalities in (9) can be concisely written in vector form as ( $\Upsilon$  is defined in (5)):  $\mathbf{e}_{\Phi} \leq \Upsilon \mathbf{e} + \mathbf{e}_{\epsilon}$ . Applying vector weighted maximum norm to this inequality results in:

$$\begin{aligned} \|\mathbf{e}_{\Phi}\|_{\infty,vec}^{\mathbf{w}} &\leq \|\Upsilon\|_{\infty,mat}^{\mathbf{w}} \|\mathbf{e}\|_{\infty,vec}^{\mathbf{w}} + \|\mathbf{e}_{\epsilon}\|_{\infty,vec}^{\mathbf{w}} \\ &= \|\Upsilon\|_{\infty,mat}^{\mathbf{w}} \|\mathbf{p}^1 - \mathbf{p}^2\|_{2,block}^{\mathbf{w}} + \|\epsilon\|_{2,block}^{\mathbf{w}}. \end{aligned} \quad (13)$$

Arguing similarly as the derivation of the Proposition 2 of [4] by using (12) and (13) we have:

$$\begin{aligned} &\|\hat{\Phi}(\mathbf{p}^1) - \Phi(\mathbf{p}^2)\|_{2,block}^{\mathbf{w}} \\ &= \|\mathbf{e}_{\Phi}\|_{\infty,vec}^{\mathbf{w}} \leq \|\Upsilon\|_{\infty,mat}^{\mathbf{w}} \|\mathbf{p}^1 - \mathbf{p}^2\|_{2,block}^{\mathbf{w}} + \|\epsilon\|_{2,block}^{\mathbf{w}}. \end{aligned} \quad (14)$$

Since  $\Upsilon$  is a non-negative matrix, from [15] Corollary 6.1, we have that there exists a  $\bar{\mathbf{w}}$  such that  $\rho(\Upsilon) < 1 \iff \|\Upsilon\|_{\infty,mat}^{\bar{\mathbf{w}}} < 1$ . Consequently, we conclude that if  $\rho(\Upsilon) < 1$ , then there must exists a  $\beta \in (0, 1)$  and a positive vector  $\bar{\mathbf{w}}$  that satisfy (6).  $\blacksquare$

## APPENDIX B

### PROOF OF THEOREM 1

*Proof:* Starting from an arbitrary initial point  $\mathbf{p}^0 \in \mathcal{P}$ , the magnitude of the difference between  $\mathbf{p}^1$  and the fixed point  $\mathbf{p}^*$  can be expressed as:

$$\begin{aligned} \|\mathbf{p}^1 - \mathbf{p}^*\| &= \|(1 - \alpha_0)\mathbf{p}^0 + \alpha_0\hat{\Phi}(\mathbf{p}^0) - \mathbf{p}^*\| \\ &\leq \|(1 - \alpha_0)(\mathbf{p}^0 - \mathbf{p}^*)\| + \left\| \alpha_0 \left( \hat{\Phi}(\mathbf{p}^0) - \Phi(\mathbf{p}^*) \right) \right\| \\ &\stackrel{(i)}{\leq} (1 - \alpha_0)\|\mathbf{p}^0 - \mathbf{p}^*\| + \alpha_0\beta\|\mathbf{p}^0 - \mathbf{p}^*\| + \|\alpha_0\epsilon^0\| \\ &= \left( 1 - \alpha_0(1 - \beta) + \frac{\|\alpha_0\epsilon^0\|}{\|\mathbf{p}^0 - \mathbf{p}^*\|} \right) \|\mathbf{p}^0 - \mathbf{p}^*\| \end{aligned} \quad (15)$$

where (i) is from Lemma 1. Let us denote  $\mu_t \triangleq (1 - \alpha_t(1 - \beta))$ . From (3) and (4), clearly we have  $\alpha_0 = 1$  and  $\alpha_t \leq 1$ , which implies  $\mu_0 \leq \mu_t \forall t$ . By induction, we show that in general:

$$\|\mathbf{p}^T - \mathbf{p}^*\| \leq \left( \prod_{t=0}^{T-1} \mu_t + \frac{\sum_{t=0}^{T-1} \left( \prod_{j=t}^{T-1} \mu_j \right) \|\alpha_t \boldsymbol{\epsilon}^t\|}{\mu_0 \|\mathbf{p}^0 - \mathbf{p}^*\|} \right) \|\mathbf{p}^0 - \mathbf{p}^*\|. \quad (16)$$

Clearly from (15) at time  $T = 1$ , (16) is true. Suppose at time  $T$ , (16) is true. At time  $T + 1$ , we have:

$$\begin{aligned} \|\mathbf{p}^{T+1} - \mathbf{p}^*\| &\leq (1 - \alpha_T) \|\mathbf{p}^T - \mathbf{p}^*\| + \alpha_T \beta \|\mathbf{p}^T - \mathbf{p}^*\| + \|\alpha_T \boldsymbol{\epsilon}^T\| \\ &\leq \left( \prod_{t=0}^T \mu_t + \frac{\sum_{t=0}^{T-1} \left( \prod_{j=t}^T \mu_j \right) \|\alpha_t \boldsymbol{\epsilon}^t\|}{\mu_0 \|\mathbf{p}^0 - \mathbf{p}^*\|} + \frac{\|\alpha_T \boldsymbol{\epsilon}^T\|}{\|\mathbf{p}^0 - \mathbf{p}^*\|} \right) \|\mathbf{p}^0 - \mathbf{p}^*\| \\ &\leq \left( \prod_{t=0}^T \mu_t + \frac{\sum_{t=0}^T \left( \prod_{j=t}^T \mu_j \right) \|\alpha_t \boldsymbol{\epsilon}^t\|}{\mu_0 \|\mathbf{p}^0 - \mathbf{p}^*\|} \right) \|\mathbf{p}^0 - \mathbf{p}^*\|. \end{aligned} \quad (17)$$

Note that in the last inequality, we have used the fact that  $\mu_t \geq \mu_0$ , and  $\frac{\|\alpha_T \boldsymbol{\epsilon}^T\|}{\|\mathbf{p}^0 - \mathbf{p}^*\|} \leq \frac{\mu_T \|\alpha_T \boldsymbol{\epsilon}^T\|}{\mu_0 \|\mathbf{p}^0 - \mathbf{p}^*\|}$ . From the assumption  $\sum_{t=1}^{\infty} \alpha_t \|\boldsymbol{\epsilon}^t\| < \infty$ , w. p. 1, there must exist some constant  $0 < b < \infty$  such that:

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T \|\alpha_t \boldsymbol{\epsilon}^t\| \leq b < \infty \text{ w. p. 1.} \quad (18)$$

In the following, we show  $\lim_{T \rightarrow \infty} \sum_{t=0}^{T-1} \left( \prod_{j=t}^{T-1} \mu_j \right) \|\alpha_t \boldsymbol{\epsilon}^t\| = 0$  w. p. 1.

First note that we have  $\lim_{T \rightarrow \infty} \prod_{t=0}^T \mu_t = 0$ , because:

$$\begin{aligned} \lim_{T \rightarrow \infty} \log \left( \prod_{t=0}^T \mu_t \right) &= \lim_{T \rightarrow \infty} \sum_{t=0}^T \log \left( 1 + (-\alpha_t(1 - \beta)) \right) \stackrel{(i)}{=} \\ &\leq \lim_{T \rightarrow \infty} (1 - \beta) \sum_{t=0}^T -\alpha_t \stackrel{(ii)}{=} -\infty \end{aligned} \quad (19)$$

where (i) is because  $-1 < -\alpha_t(1 - \beta)$  and the fact  $\log(1 + x) \leq x, \forall x > -1$ , (ii) is because (4) and  $\beta < 1$ . Clearly (19) implies  $\lim_{T \rightarrow \infty} \prod_{t=0}^T \mu_t = 0$ . Thus for any  $\delta > 0$ , and a fixed  $T$  there exists  $\hat{T}(T, \delta) > T$  such that:

$$\prod_{t=T}^{N-1} \mu_t \leq \frac{\delta}{2b}, \forall N \geq \hat{T}(T, \delta). \quad (20)$$

From (18) we have that for any  $\delta > 0$ , there exists  $T(\delta)$  such that:

$$\sum_{t=T}^{\infty} \|\alpha_t \boldsymbol{\epsilon}^t\| \leq \frac{\delta}{2}, \forall T \geq T(\delta), \text{ w. p. 1.} \quad (21)$$

Then we have that for all  $N > \max \{T(\delta), \hat{T}(T(\delta), \delta)\} = \hat{T}(T(\delta), \delta)$ :

$$\begin{aligned}
& \sum_{t=0}^N \left( \prod_{j=t}^{N-1} \mu_j \right) \|\alpha_t \epsilon^t\| \\
&= \sum_{t=0}^{T(\delta)} \left( \prod_{j=t}^{N-1} \mu_j \right) \|\alpha_t \epsilon^t\| + \sum_{t=T(\delta)+1}^N \left( \prod_{j=t}^{N-1} \mu_j \right) \|\alpha_t \epsilon^t\| \\
&\stackrel{(i)}{\leq} \sum_{t=0}^{T(\delta)} \left( \prod_{j=T(\delta)}^{N-1} \mu_j \right) \|\alpha_t \epsilon^t\| + \frac{\delta}{2} \\
&\stackrel{(ii)}{\leq} \prod_{j=T(\delta)}^{N-1} \mu_j \sum_{t=0}^{T(\delta)} \|\alpha_t \epsilon^t\| + \frac{\delta}{2} \stackrel{(iii)}{=} \\
&\leq \frac{\delta}{2b} b + \frac{\delta}{2} = \delta \quad \text{w. p. 1}
\end{aligned} \tag{22}$$

where (i) is because (21) and the fact that  $\prod_{j=t}^{N-1} \mu_j < 1$  for all  $t \leq N-1$ ; (ii) is because  $\prod_{j=T(\delta)}^{N-1} \mu_j$  is independent of  $t$ ; (iii) is because of (18) and (20). Consequently, we have that:

$$\lim_{T \rightarrow \infty} \sum_{t=0}^{T-1} \left( \prod_{j=t}^{T-1} \mu_j \right) \|\alpha_t \epsilon^t\| = 0 \quad \text{w. p. 1.} \tag{23}$$

From (16), (23), and  $\lim_{T \rightarrow \infty} \prod_{t=0}^T \mu_t = 0$ , we conclude :  $\lim_{t \rightarrow \infty} \|\mathbf{p}^t - \mathbf{p}^*\| = 0$  w. p. 1. ■

## APPENDIX C

### PROOF OF THEOREM 2

*Proof:* Due to space limit, we only show the proof for the case that  $\alpha_t = \frac{1}{t+1}$ . The proof for general  $\{\alpha_t\}$  can be obtained similarly. When taking  $\alpha_t = \frac{1}{t+1}$ , the A-IWF algorithm can be written compactly as:  $\mathbf{p}^{T+1} = \frac{1}{T+1} \sum_{t=0}^T \hat{\Phi}(\mathbf{p}^t)$ . We can write:

$$\begin{aligned}
\|\mathbf{p}^{T+1} - \mathbf{p}^*\| &\leq \frac{1}{T+1} \sum_{t=0}^T \|\hat{\Phi}(\mathbf{p}^t) - \Phi(\mathbf{p}^*)\| \\
&\stackrel{(i)}{\leq} \frac{\beta}{T+1} \sum_{t=0}^T \|\mathbf{p}^t - \mathbf{p}^*\| + \frac{1}{T+1} \sum_{t=0}^T \|\epsilon^t\|
\end{aligned} \tag{24}$$

where (i) is from Lemma 1. Suppose the sequence  $\{\mathbf{p}^t\}$  does not converge to  $\mathbf{p}^*$ , i.e.,  $\lim_{T \rightarrow \infty} \sup \|\mathbf{p}^T - \mathbf{p}^*\| = \delta > 0$ . Using the Stolz-Cesàro Theorem [17], we have that:

$$\begin{aligned}
\lim_{T \rightarrow \infty} \sup \frac{\sum_{t=0}^T \|\mathbf{p}^t - \mathbf{p}^*\|}{T+1} &\leq \lim_{T \rightarrow \infty} \sup \|\mathbf{p}^T - \mathbf{p}^*\| = \delta; \\
\lim_{T \rightarrow \infty} \frac{\sum_{t=0}^T \|\epsilon^t\|}{T+1} &= \lim_{T \rightarrow \infty} \|\epsilon^T\| = 0, \quad \text{w. p. 1.}
\end{aligned} \tag{25}$$

Taking  $\limsup$  on both sides of (24), we have:

$$\begin{aligned} & \lim_{T \rightarrow \infty} \sup \|\mathbf{p}^{T+1} - \mathbf{p}^*\| \\ & \leq \lim_{T \rightarrow \infty} \sup \frac{\beta}{T+1} \sum_{t=0}^T \|\mathbf{p}^t - \mathbf{p}^*\| + \lim_{T \rightarrow \infty} \sup \frac{1}{T+1} \sum_{t=0}^T \|\epsilon^t\| \end{aligned} \quad (26)$$

which can be reduced to:  $\delta \leq \beta\delta$  by applying (25). This is a contradiction to the fact that  $\beta < 1$ . Then we conclude that  $\lim_{T \rightarrow \infty} \sup \|\mathbf{p}^T - \mathbf{p}^*\| = 0$  which in turn implies  $\lim_{T \rightarrow \infty} \|\mathbf{p}^T - \mathbf{p}^*\| = 0$ . ■

#### APPENDIX D

##### PROOF OF THEOREM 3

Due to space limit, we only show the proof for the case that  $\alpha_t = \frac{1}{t+1}$ . The proof for general  $\{\alpha_t\}$  can be obtained similarly. We first state a lemma, the proof of which can be found in Appendix E.

*Lemma 2:* If  $w^{t+1} = (1 - \alpha_t)w^t + \alpha_t \xi^{t+1}$ , and  $E[\xi^t | \mathcal{F}^{t-1}] = 0$ , and  $E[(\xi^t)^2 | \mathcal{F}^{t-1}] = b$  is uniformly bounded,  $\{\alpha_t\}$  satisfies (4), then we must have  $\lim_{t \rightarrow \infty} w^t = 0$ , w. p. 1.

We are now ready to prove Theorem 3. The A-IWF algorithm can be compactly written as:  $\mathbf{p}^{T+1} = \frac{1}{T+1} \sum_{t=0}^T \hat{\Phi}(\mathbf{p}^t) = \frac{1}{T+1} \sum_{t=0}^T \Phi(\mathbf{p}^t) + \mathbf{w}^T$ , where  $\mathbf{w}^T \triangleq \frac{1}{T+1} \sum_{t=0}^T \xi^t = (1 - \frac{1}{T+1})\mathbf{w}^{T-1} + \frac{1}{T+1}\xi^T$ . Note that by applying the results of Lemma 2, we have  $\lim_{T \rightarrow \infty} \mathbf{w}^T = 0$ . Then the magnitude of difference between  $\mathbf{p}^{T+1}$  and the unique fixed point of the mapping  $\Phi(\cdot)$  can be expressed as:

$$\begin{aligned} \|\mathbf{p}^{T+1} - \mathbf{p}^*\| & \leq \frac{1}{T+1} \sum_{t=0}^T \|\Phi(\mathbf{p}^t) - \Phi(\mathbf{p}^*)\| + \|\mathbf{w}^T\| \\ & \leq \frac{\beta}{T+1} \sum_{t=0}^T \|\mathbf{p}^t - \mathbf{p}^*\| + \|\mathbf{w}^T\|. \end{aligned} \quad (27)$$

Suppose the sequence  $\{\mathbf{p}^t\}$  does not converge to  $\mathbf{p}^*$ , then there must exist a  $\delta > 0$  such that  $\lim_{T \rightarrow \infty} \sup \|\mathbf{p}^t - \mathbf{p}^*\| = \delta$ . Using again the Stolz-Cesàro Theorem as in (25), and taking  $\limsup$  on both sides of (27), we have:  $\lim_{T \rightarrow \infty} \sup \|\mathbf{p}^{T+1} - \mathbf{p}^*\| \leq \lim_{T \rightarrow \infty} \sup \frac{\beta}{T+1} \sum_{t=0}^T \|\mathbf{p}^t - \mathbf{p}^*\| + \lim_{T \rightarrow \infty} \|\mathbf{w}^T\|$ . This inequality can be reduced to:  $\delta \leq \beta\delta$ , which contradicts to the fact that  $\beta < 1$ . Thus we conclude that  $\lim_{T \rightarrow \infty} \|\mathbf{p}^t - \mathbf{p}^*\| = 0$ , and that  $\lim_{T \rightarrow \infty} \mathbf{p}^t = \mathbf{p}^*$ .

#### APPENDIX E

##### PROOF OF LEMMA 2

*Proof:* We have  $w^{t+1} = w^t + \alpha_t(\xi^{t+1} - w^t)$ . Consider the following iteration:

$$\begin{aligned} (w^{t+1})^2 & = (w^t + \alpha_t(\xi^{t+1} - w^t))^2 \\ & = (w^t)^2 + 2\alpha_t(\xi^{t+1} - w^t)w^t + \alpha_t^2(\xi^{t+1} - w^t)^2. \end{aligned} \quad (28)$$



Then  $E[(w^{t+1})^2|\mathcal{F}^t]$  can be expressed as:

$$\begin{aligned}
& E[(w^{t+1})^2|\mathcal{F}^t] \\
&= (w^t)^2 - 2\alpha_t(w^t)^2 + 2\alpha_t E[\xi^{t+1}|\mathcal{F}^t]w^t + \alpha_t^2 E[(\xi^{t+1} - w^t)^2|\mathcal{F}^t] \\
&= (w^t)^2 - 2\alpha_t(w^t)^2 + \alpha_t^2 (E[(\xi^{t+1})^2|\mathcal{F}^t] + (w^t)^2 - 2w^t E[\xi^{t+1}|\mathcal{F}^t]) \\
&\leq (w^t)^2 - 2\alpha_t(1 - \frac{\alpha_t}{2})(w^t)^2 + \alpha_t^2 b.
\end{aligned} \tag{29}$$

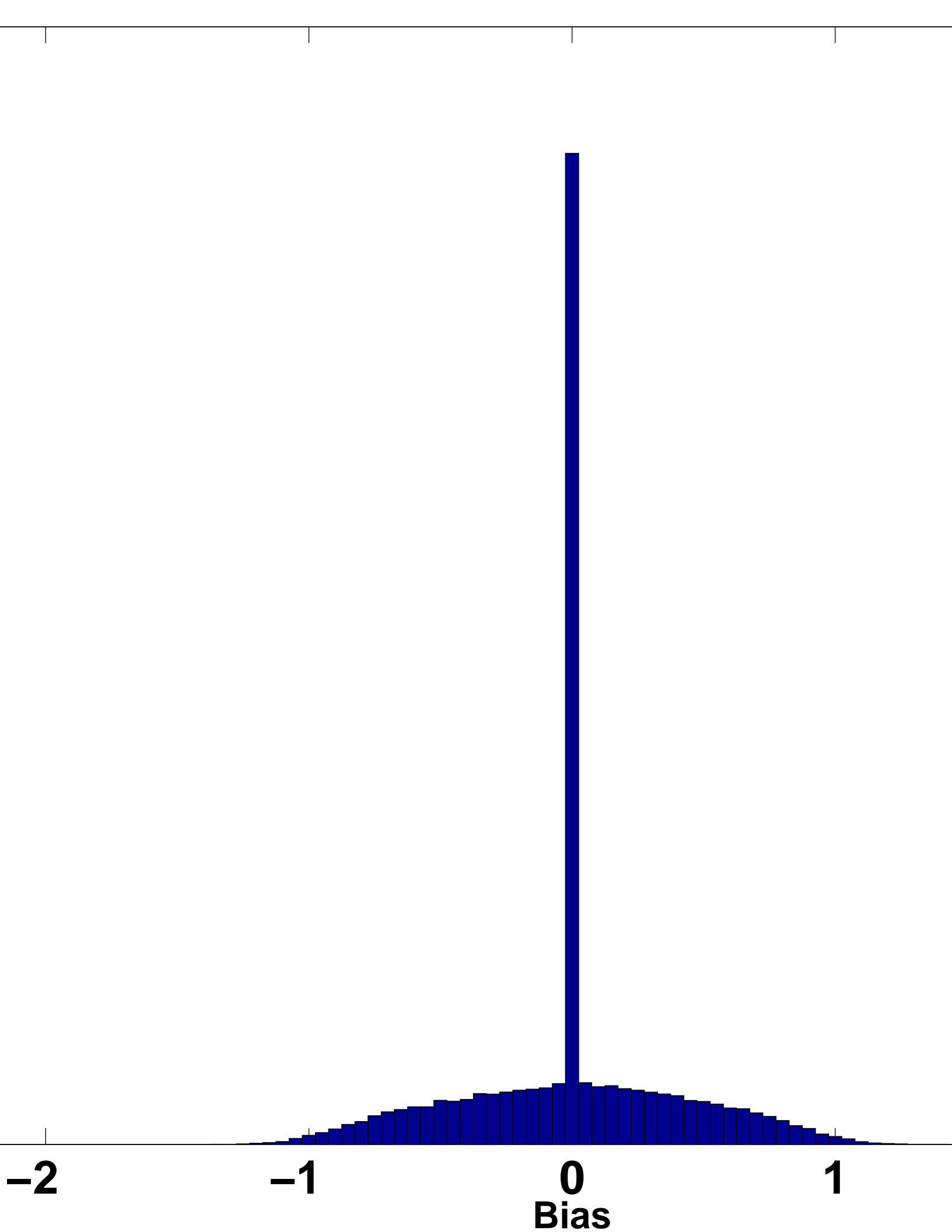
Notice that the term  $2\alpha_t(1 - \frac{\alpha_t}{2})(w^t)^2 \geq 0$  because  $0 < \alpha_t \leq 1$ . We see that  $\lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t^2 b < \infty$  because  $\sum_{t=1}^{\infty} \alpha_t^2 < \infty$ . In order to proceed, we define the notion of a *non-negative almost-supermartingale* [18]. Let  $z_t$ ,  $\beta_t$ ,  $\xi_t$  and  $\zeta_t$  be *non-negative*  $\mathcal{F}^t$  measurable random variables. The sequence  $\{z_t\}$  is called *non-negative almost-supermartingale* if  $E[z_{t+1}|\mathcal{F}^t] \leq (1 + \beta_t)z_t + \xi_t - \zeta_t$ . From Theorem 1 of [18], we have  $\lim_{t \rightarrow \infty} z_t$  exists and is finite and  $\sum_{t=1}^{\infty} \zeta_t < \infty$  w. p. 1 if  $\{\sum_{t=1}^{\infty} \beta_t < \infty, \sum_{t=1}^{\infty} \xi_t < \infty\}$ .

Now it is clear that the sequence  $\{(w^t)^2\}_{t=0}^{\infty}$  is a non-negative almost-supermartingale, and according to the above mentioned theorem we have the following results: 1)  $\{(w^t)^2\}_{t=0}^{\infty}$  converges; 2)  $\sum_{t=1}^T \alpha_t(1 - \frac{\alpha_t}{2})(w^t)^2 < \infty$  w. p. 1. The second result implies that  $\lim_{T \rightarrow \infty} \sum_{t=1}^T \alpha_t(w^t)^2 < \infty$ . Combined with the fact that  $\sum_{t=0}^{\infty} \alpha_t = \infty$  and  $\lim_{t \rightarrow \infty} \alpha_t = 0$ , we have that  $\liminf_{t \rightarrow \infty} (w^t)^2 = 0$ . Moreover, we know from the first result that the sequence  $\{(w^t)^2\}_{t=0}^{\infty}$  converges, then it must converge to 0. ■

## REFERENCES

- [1] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, 2002.
- [2] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory – part I: Nash equilibria," *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1230–1249, 2008.
- [3] F. Wang, M. Krunz, and S. G. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 74–87, 2008.
- [4] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory – part II: Algorithms," *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1250–1267, 2008.
- [5] G. Scutari, D. P. Palomar, and S. Barbarossa, "Asynchronous iterative water-filling for Gaussian frequency-selective interference channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2868–2878, 2008.
- [6] Z.-Q. Luo and J.-S. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–10, 2006.
- [7] K. W. Shum, K. K. Leung, and C. W. Sung, "Convergence of iterative waterfilling algorithm for Gaussian interference channels," *IEEE Journal on Selected Area in Communications*, vol. 25, pp. 1091–1100, 2007.
- [8] A. Leshem and E. Zehavi, "Game theory and the frequency selective interference channel," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 28–40, 2009.

- [9] T. R. Benedict and T. T. Soong, "The joint estimation of signal and noise from the sum envelope," *IEEE Transactions on Information Theory*, vol. 13, no. 3, pp. 447–454, 1967.
- [10] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Transactions on Communications*, vol. 48, no. 10, pp. 1681–1691, 2000.
- [11] P. Setoodeh and S. Haykin, "Robust transmit powercontrol for cognitive radio," *Proceedings of IEEE*, pp. 915–939, 2009.
- [12] R. H. Gohary and T. J. Willink, "Robust IWFA for open-spectrum communications," *IEEE Trans. On Signal Process.*, vol. 57, no. 12, pp. 4964–4970, 2009.
- [13] Y. Cheng and V. K. N. Lau, "Distributive power control algorithm for multicarrier interference network over time-varying fading channels—tracking performance analysis and optimization," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4750–4760, 2010.
- [14] W. R. Mann, "Mean value methods in iteration," in *Proc. Amer. Math.Soc.*, 1953, pp. 506–510.
- [15] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, 1997.
- [16] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1972.
- [17] M. Muresan, *A Concrete Approach to Classical Analysis*, Springer, 2008.
- [18] H. Robbins and D. Siegmund, *A Convergence Theorem for Non-Negative Almost Supermartingales and Some Applications*, Optimizing Methods in Statistics. Academic Press, New York, 1971.



## Comparison of Convergence of Different Algorithms

